An Expert System for Psychiatric Diagnosis using the DSM-III-R, DSM-IV and ICD-10 Classifications

Roland H.C. Yap[1]   David M. Clarke[2]

[1] Dept. of Information Systems & Computer Science, National University of Singapore, Lower Kent Ridge Road, Singapore 119260, Singapore.

[2] Monash University Department of Psychological Medicine, Monash Medical Centre, 246 Clayton Road, Clayton 3168, Australia.

Making a systematic and comprehensive psychiatric evaluation of mental disorders in a patient can be a rather complex and involving process. We describe an expert system, MILP, which is designed to produce such systematic diagnoses of mental disorders using selected categories from the classification and diagnostic guidelines published in DSM-III-R, DSM-IV and ICD-10. An innovative part of the MILP design is the incorporation of constraint-based reasoning as a key part of the system. We believe that the MILP design gives a flexible framework which is suitable in general for the automated diagnoses of large classes of mental disorders.

INTRODUCTION

In this paper, we consider computational techniques and frameworks which are suitable for psychiatric diagnosis. In particular, we are studying methods for the systematic diagnosis and characterization of mental disorders in patients. This is a difficult task because the definition of a mental disorder, by its very nature, is often not completely precise. To systematize the task of diagnosis, there are a number of "standard" classifications and definitions of mental disorders. These are the DSM-III-R (Diagnostic and Statistical Manual of Mental Disorders) [1], produced by the American Psychiatric Association, which was later revised to the DSM-IV classification [2], and the ICD-10 (International Classification of Diseases) [3,4], produced by the World Health Organization.

We first began this work because the second author had designed a structured interview based on the DSM-III-R, DSM-IV and ICD-10 criteria for the purpose of performing systematic diagnosis studies. This interview consisted of a questionnaire form of about 38 pages in length. The interview was used to collect patient data using a manual interview process lasting on the order of half an hour per patient interviewed. A large number of patients (300) were studied to collect all the interview data. This interview database thus forms a rich source of data suitable for several kinds of psychiatric studies. With the amounts of data collected and the number of diagnoses which would be needed for any kind of comprehensive study, it was clear that an effective computational technique was a necessity.

Sophisticated systems for medical diagnoses using techniques from expert systems abound in the literature. Two well-known examples are MYCIN [5] for the diagnosis of infectious diseases and INTERNIST/CADUCEUS [6] for internal medicine (see also [7] for a text). In the area of psychiatric diagnoses, we are not aware of computer programs of comparable sophistication. The kinds of programs which are available are programs such as DTREE[8] and AUTOSCID. DTREE is an expert system for diagnosing DSM-IV (old version has DSM-III-R) Axis I disorders. AUTOSCID is a computerized version of the Structured Clinical interview for DSM-IV Axis II personality disorders. Both these systems are essentially interactive programs for diagnostic help using the DSM-IV criteria. For example, DTREE, appears to be based on simple "decision tree" techniques.

The goals of our original study was to have a system capable of diagnosing using many classification schemes such as DSM-III-R, DSM-IV, ICD-10 and also various other miscellaneous criteria. Thus the scope of the study goes beyond the offerings of programs like DTREE and AUTOSCID. Another important goal was to ensure that the definitions for the diagnostic criteria (rules) be flexible and "easily" modifiable. This turned out to be

fairly important as constant changes were made in the refining of the diagnostic rules. The database for the interview is different from the kind of input needed for programs like DTREE, since the interview is not only different but also deals with a larger classification base. Because of these requirements, we built our own expert system, called MILP (Monash Interview for Liaison Psychiatry).

DESIGN AND METHODOLOGY

The MILP expert system which we have developed is written in the Constraint Logic Programming (CLP) language called CLP($\mathcal{R}$) [9]. Constraint logic programming [10] is a framework for incorporating constraints into logic programming based languages. Complex problems are often complex partly because of the "problem constraints" which are the inter-relationships between various variables and states within the problem. CLP languages provide a direct way of capturing/modelling these relationships using constraints and thus offer greater expressive power and flexibility. It is for this reason, that there are a number of model-based reasoning applications developed using CLP. One medical reasoning application is the work of De Geus et al [11] where CLP is used to model human circulation and gas exchange using constraints to specify the fluid flow.

Logic programming languages, such as Prolog, have also been popular in building expert systems because of the close relationship between expert system rules and logical deduction, see [12] for a survey on applying logic-based techniques to medicine. There is also some recent work on incorporating ideas from CLP more directly into expert systems [13].

The MILP system was written in CLP($\mathcal{R}$) in order to take advantage of the constraint and logic programming aspects of CLP($\mathcal{R}$). The logic programming (Prolog) and constraint features turn out to be both convenient and also point to more interesting uses for such expert systems.

Problem Scope

Here we describe some of the main design criteria which we required for the initial study. Note that while only some diagnostic categories were used, we will argue that the system design is suitable for other diagnoses as well. The kinds of diagnoses

which we considered are essentially those diagnoses which are amenable to diagnosis by a structured interview process. It is particularly comprehensive in the areas of anxiety, depression and somatoform disorders. Schizophrenia and other psychoses were not dealt with as they are uncommon in the particular patient population under consideration. The following are categories of disorders considered, together with some examples, which are based on the criteria and classification from DSM-III-R, DSM-IV and ICD-10:

- Mood disorders — manic and hypomanic episode, major depression, etc.

- Anxiety disorders — panic disorder, social phobia, obsessive compulsive disorder, etc.

- Somatic Disorders — somatization disorder, hypochondriasis, neurasthenia, etc.

- Substance dependence & harmful use — alchohol, sedative, nicotine, etc.

- Eating disorders — anorexia nervosa and bulimia nervosa.

Not every disorder from the above will occur in the various classifications and in some cases the language and specifics also differ. In addition, we had other miscellaneous categories and classifications, for example: grief, psychotic symptoms, stewart depression, etc. Note that some of the diagnoses are not the final diagnoses and are only steps on the diagnostic pathway. Apart from the diagnoses themselves, each diagnoses also may have various specifiers such as currency, acute/chronic, etc.

The interview dataset after encoding into a flat form is quite large. The rules for the various disorders in total require approximately 600 data items per patient. The exact number needed depends on the precise rule definitions. This amount is listed to give an idea of the amount of data which is used to obtain a single comprehensive patient diagnosis. The data itself consists of boolean, numeric or item selections.

The MILP system

The MILP system is designed around a toolkit philosophy. It consists of a generic custom inference engine, a front-end program to the interview database, an output module for human readable diagnoses and an encoding module. A possible run

of the program will involve using the inference engine with the appropriate rule database, say ICD-10. The interview database is read (this can be viewed as some input in Prolog-style facts/terms), all the appropriate diagnoses are generated for a selected rule database, and the generated diagnoses are also Prolog-style facts/terms. The generated diagnoses can then either be output in human readable form by the output module (omitting intermediate diagnoses if desired) or encoded using the encoding module for subsequent statistical analysis by SPSS.

The built-in database aspects of Prolog in CLP($\mathcal{R}$) provide a convenient self-describing system. The expert rules, descriptions and explanations of diagnoses, encoding of output, etc. are all part of the database as a combination of Prolog rules, MILP rules and table-driven data.

The Expert Language

Since we are using a custom inference engine, there is freedom to choose the language to be used by the human expert in which to write the rules. For reasons of space, we cannot describe the expert language (henceforth, simply as language) in detail and will illustrate it by way of example. The main design goals for the language are expressiveness – can it express all the kinds of rules which we would like to write in a natural fashion? For example, since many of the constraints are boolean expressions, they could all be written in disjunctive normal form. However, that would not only be un-natural and non-declarative, but also textually large and unwieldy. For example, consider writing a condition like: *at least five out of ten sub-conditions must be satisfiable* into normal form. We have chosen to represent the rules in two highly related languages. One is a semi-formal language, *expert language*, for the human expert and the other is the actual formal *rule language* used by the system. However, these two forms are very close to each other and in most cases, a simple one-to-one mapping suffices to do the conversion. The expert language is meant to be understandable by doctors who are familiar with the disorder in question, and also understand the underlying questions from the interview, eg. the question MAN1 in the following example. An example of the semi-formal rule is as follows (it has been simplified by omitting some conditions to make the example more compact):

ICD-10 Hypomania — Specify: currency

A. *Elevated/* MAN1 or MAN2
expansive or irritable mood
B. *At least 3 of:*
1. Increased activity/restlessness
MAN8 or MAN9
2. Increased talkativeness
MAN4
3. Decreased concentration
MAN7
. . . . . .
7. Increased sociability
MAN5
C. *Not organic* MAN14 = 5,6,7 or 8
D. *Exclusions:* Mania
E. *Currency:*
(Current) MAN16 = 1 or 2
(Not current) otherwise

Data items are those in uppercase like MAN1 above. Hypomania is defined as having a currency specifier (current or not-current). Various types of conditions can be specified, in condition **B**, the rule states that at least three of the seven sub-conditions must be satisfied and those in turn can, in general, be any arbitrarily nested condition. This is then translated in a straightforward fashion into the following rule language used by the system. Note the intentional similarity to a Prolog clause. A Prolog-style term syntax has again been adopted intentionally to make bracketing and nesting clear.

```
icd(hypomania) :-
    or([man1, man2]),          % A
    atleast(3, [               % B
        or([man8, man9]),
        man4, man7, ...,
        man5
    ]),
    man14 = values([5,6,7,8]), % C
    notprovable(mania),        % D
    setcurrent(                % E
        man16 = values([1,2]), Currency
    ),
    assert(hypomania, [Currency]).
```

A rule contains boolean expressions which are implicitly conjoined together with other constraint relations (which can also contain arithmetic expressions) as well as some special actions. The exclusion condition is written as `notprovable(mania)` which specifies the requirement that this diagnosis is not consistent with mania and hence to diagnose hypomania this rule,

231

we must ensure that mania is not provable by the inference engine. Specifiers can also be added to exclusions which means that a diagnosis is then only excluded if it can be proved to have at least those specifiers. `Setcurrent` is a special action which sets `Currency` to be current or not-current depending on whether the first argument is satisfiable or not. Finally, `assert(hypomania, [Currency])`, denotes that if the above conditions are satisfiable then we can diagnose ICD-10 hypomania and it has one specifier given in the variable `Currency`. Note that unification (equality constraints) can link the subparts of the rule (as with `Currency`).

The Inference Engine

The inference engine is a meta-interpreter (see [14]) for the rule language above using a backward-chaining mechanism. Arithmetic constraints are translated directly into CLP($\mathcal{R}$) constraints. Boolean constraints are simulated by encoding into arithmetic constraints. For example, the boolean constraint $X = \wedge_{i=1}^{n} Y_i$ can be partially simulated with arithmetic constraints as

$$X \leq Y_i \leq 1, 0 \leq X \leq 1, \sum_{i=1}^{n} Y_i \leq X + n - 1.$$

A CLP system with native constraints which can express all the kinds of arithmetic and boolean constraints required by the language would be better as constraints can then be solved more directly.

ASSESSMENT OF RESULTS

In turned out that the design used was quite simple to implement but yet also flexible. A substantial portion of the time was spent in debugging and refining of the rules themselves. Occasionally, the rule language was extended if there was a need but this was quite a straightforward process. The MILP system is the result of an interaction between understanding what rules need to be expressed and development of an appropriate inference engine which understands the interpretation of those rules.

Codifying the knowledge contained in DSM-III-R, DSM-IV and ICD-10 was a major task. In many cases, substantial modifications to a large number of rules were needed because of changes in the interpretation of particular diagnoses. Without a flexible, expressive rule language tailored to

this task, it would have been much less feasible for the system to undergo constant change.

Verifying the result is also a difficult problem. How confident is one that the rules codify the knowledge and that the diagnoses are both complete (in the comprehensive sense) and correct? We used essentially two types of validation strategies. One was random sampling of patients to rigorously check the result against how an expert would interpret them. It is usually not possible to do this kind of testing in an exhaustive fashion.

Another strategy is to make use of the fact that there is strong correlation and commonality between DSM-III-R, DSM-IV and ICD-10. While there are sufficient systematic differences between the three diagnoses schemas, one can still check the correlation between the diagnosis in a number of ways. One way is to have a use a thorough check of substantial differences when they occur. Another is to look at statistical correlation (see [15] for an analysis of the results) between classes of diagnoses either inter-schema or intra-schema.

Status

The existing MILP program has been used for a cross validation study (analogous to the validation of the results themselves above) of the MILP interview [16]. It was then used for a study of breast cancer patients which monitors the patients over their treatment period. It is interesting to observe that in a number of cases of the breast cancer study, that the program turned up additional diagnoses which were initially missed since the program produces comprehensive diagnoses.

Existing work is now in progress to interface the inference engine and associated modules and rule database to a front-end graphical user interface. This will allow the program to be easily used as a stand-alone diagnosis tool and the intent is to trial it out in a number of sites around Australia.

DISCUSSION

We have used a constraint-based approach to building the expert system. This was done partly because we knew early in the design that a flexible and expressive language and system was going to be a key requirement. It turned out that using constraints meant that the development of the inference engine and rule-bases was much easier as portions of it could be tested with incomplete

data. The constraint features of the design was used mostly during the system development but was not exploited for the generation of the final patient diagnoses because all the required data items were complete. Direct use of the MILP for generating diagnoses thus benefits from the constraint design but does not require it. However other uses of the MILP described below can take advantage of constraints.

From a medical research viewpoint, the MILP system combined with the interview database can be a useful tool for all kinds of (medical) studies. Apart from the obvious direct uses of the system as a diagnosis generation tool, one can also study the differences between diagnoses generated with respect to DSM-III-R, DSM-IV and ICD-10. It can also be used as a teaching tool. Furthermore, the rule language used in the MILP seems to capture the main kinds of conditions and concepts needed for these kinds of psychiatric diagnoses. We expect that it will be easy to extend or modify the MILP for other diagnostic purposes.

In the paper by Haug [17], he calls for a broader use of diagnostic expert systems than just generating diagnoses. A system such as the MILP which is designed to use constraints for flexibility and expressiveness has many potential uses in this regard. For example, it can be used to analyse the rules themselves rather than doing diagnoses. Suppose we are interested in knowing in what context can three rules (diagnoses) hold at the same time. Now, instead of generating a diagnosis, we are asking whether the inference engine can prove the conditions of the three rules (together with any subsidiary rules and requirements) can hold at the same time, in a particular context, where we might have some assumptions or no assumptions. So various questions involving meta-reasoning about the rules might also be an interesting area of research.

Acknowledgments

References

1. *Diagnostic and statistical manual of mental disorders: DSM-III-R.* American Psychiatric Association, 1987.

2. *Diagnostic and statistical manual of mental disorders: DSM-IV.* American Psychiatric Association, 1994.

3. *The ICD-10 classification of mental and behavioral disorders: clinical descriptions and diagnostic guidelines.* World Health Organization, 1992.

4. *The ICD-10 classification of mental and behavioral disorders: diagnostic criteria for research.* World Health Organization, 1992.

5. Buchanan BG, Shortliffe EH. *Rule-based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project.* Addison-Wesley, Reading, MA, 1984.

6. Miller AM, Pople HE, Myers JD. "INTERNIST-I, an experimental computer-based diagnostic consultant for general internal medicine", *New England Journal of Medicine*, 1982:307:468–476.

7. Lucas PJF, Van der Gaag LC. *Principles of Expert Systems.* Addison-Wesley, 1991.

8. First MB, Williams JBW, Spitzer RL, *Concise Guide to Using DTREE: The Electronic DSM-III-R*, American Psychiatric Association, 1988.

9. Jaffar J, Michaylov M, Stuckey PJ, Yap RHC. "The CLP($\mathcal{R}$) Language and System", *ACM Transactions on Programming Languages and Systems*, 1992:14(3):339–395.

10. Jaffar J, Maher M. "Constraint Logic Programming: a Survey", *Journal of Logic Programming*, 1994: 19&20:503–581.

11. De Gues F, Rotterdam E, Van Denneheuvel S, Van Emde Boas P. "Psysiological modelling using RL", In: Stefanelli M, Hasman A, Fieschi M, Talmon J (eds) *AIME91: Lecture Notes in Medical Informatics*, 198–210. Springer Verlag, 1991.

12. Lucas PJF. "Logic engineering in medicine", *The Knowledge Engineering Review*, 1995:10(2):153–179.

13. Liu B, Jaffar J. "Using Constraints to Model Disjunctions in Rule-Based Reasoning", To appear in *Thirteenth National Conference on Artificial Intelligence (AAAI-96)*.

14. Sterling L, Shapiro E. *The Art of Prolog.* MIT Press, 2nd ed., 1994.

15. Clark DM, Smith GC, Herman HE, McKenzie DP. "The Monash Interview for Liaison Psychiatry (MILP): development, reliability and validity", in preparation.

16. Clarke DM, Smith GC, Herman HE, McKenzie DP. "The Monash Interview for Liaison Psychiatry" (abstract), *Australian Society for Psychiatric Research*, 1995.

17. Haug PJ. "Uses of Diagnostic Expert Systems in Clinical Care", *Seventeenth Annual Symposium on Computer Applications in Medical Care: Patient-Centered Computing*, 379–383. McGraw-Hill, 1993.